A Pattern Recognition Framework to Investigate the Neural Correlates of Music

Ana Gabriela Guedes*[1]
anagabriela.a.g@hotmail.com
Alexandre Sayal*[1]
alexandresayal@gmail.com
Renato Panda[2,3]
panda@dei.uc.pt
Rui Pedro Paiva[2]
ruipedro@dei.uc.pt
Bruno Direito[1,2]
bruno.direito@uc.pt

[1] Coimbra Institute for Biomedical Imaging and Translational Research
University of Coimbra
Coimbra, Portugal

[2] Center For Informatics and Systems
University of Coimbra,
Coimbra, Portugal

[3] Ci2 — Smart Cities Research Center, Polytechnic Institute of Tomar, Tomar, Portugal

Abstract

Music can convey fundamental emotions like happiness and sadness and more intricate feelings such as tenderness or grief. Understanding the neural mechanisms underlying music-induced emotions holds promise for innovative, personalised neurorehabilitation therapies using music. Our study investigates the link between perceived emotions in music and their corresponding neural responses, measured using fMRI. Fifteen participants underwent fMRI scans while listening to 96 musical excerpts categorised into quadrants based on Russell's valence-arousal model. Neural correlates of valence and arousal were identified in neocortical regions, especially within music-specific sub-regions of the auditory cortex. Through multivariate pattern analysis, distinct emotional quadrants were decoded with an average accuracy of 62% ±15%, surpassing the chance level of 25%. This capacity to discern music's emotional qualities has implications for psychological interventions and mood modulation, enhancing music-based treatments and neurofeedback learning.

1 Introduction

The benefits of music are well-documented across various domains. Engaging in musical activities has been shown to positively affect physical rehabilitation, pain management, stress reduction, immune function, and cognitive skills enhancement. Music also represents a powerful tool for emotional regulation, enabling individuals to cope with and alleviate negative feelings such as anxiety, loneliness, and stress while cultivating positive moods such as relaxation or arousal.

Given its profound influence on human emotions and well-being, it is essential to delve into the underlying mechanisms by which music facilitates mood regulation. This includes a comprehensive understanding of the neural correlates of music-evoked emotions, which will contribute to developing neuro-rehabilitative music-based therapeutic approaches, particularly for disorders characterised by impaired emotional regulation.

Pursuing a greater understanding of human emotions has led to the proposal of different emotional theories or models, each offering unique perspectives on the comprehension of affective experiences. Russell's circumplex model of affect describes emotions in a two-dimensional plane using an unpleasantness/pleasantness dimension (valence) and a high/low arousal dimension (activation) [4]. A linear combination of these two dimensions, or different intensities of both valence and arousal, can be used to conceptualise each emotion.

Understanding how music elicits emotions combines exploring specific emotions linked to musical pieces and the underlying mechanisms through which they are evoked. Various neuroimaging techniques have been used to study the dynamic activation patterns of brain regions associated with emotions during different tasks and stimuli. Functional magnetic resonance imaging (fMRI) has become one of the most powerful techniques for evaluating brain function in clinical and research contexts, as it provides a very high resolution with whole brain coverage.

Previous meta-analyses have demonstrated that brain structures involved in music-evoked emotions are located in the limbic system, including the amygdala, hippocampus, and parahippocampal gyrus, as well as the insula, anterior cingulate cortex (ACC), and orbitofrontal cortex (OFC) [2].

The present study addresses the neural correlates of music listening by using multivariate pattern analysis (MVPA) of functional magnetic resonance imaging (fMRI) data to identify brain regions that encode emotional states. So far, most studies have focused on very restricted music stimuli (few genres and/or music excerpts) and usually use mass-univariate approaches. To this end, we used a large set of music stimuli previously labelled according to Russell's circumplex model of affect [3] and applied an MVPA approach to identify brain regions that encode emotional states.

2 Methods

2.1 Participants

Fifteen individuals (9 females, 6 male; age range 22–41 years, M=31.7, SD=6.27) participated in the experiment. All participants gave written informed consent. The study was conducted in accordance with the declaration of Helsinki and approved by the Comissão de Ética e Deontologia da Investigação da Faculdade de Psicologia e Ciências de Educação da Universidade de Coimbra. All the volunteers reported normal hearing, without permanent or current temporary impairments and with no known history of neurological disorders.

2.2 Setup and experimental protocol

Each participant listened to 96, 11.5 seconds musical excertps, randomly selected: 24 for each quadrant. Within each run, participants listened to 24 excerpts, 6 of each quadrant grouped together in blocks of 3. Inter-stimuli intervals (ISI) blocks consisted of 36 seconds (a 12-second segment of white noise, auditory control condition, interleaved with two 12-second segments of ambient noise in the ISI) (see Figure 1).

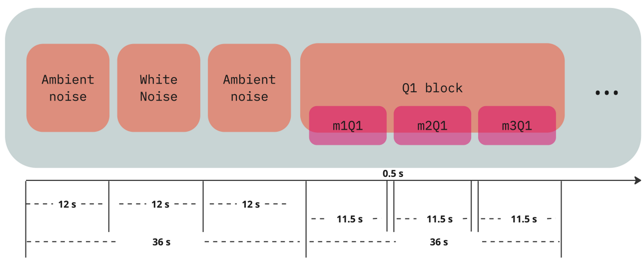

Figure 1: Experimental protocol. Each run consisted of music presentation blocks (two repetitions, four quadrants) interleaved with ISI.

MR acquisition was performed with a 3 T Siemens Magnetom Prisma scanner with a 20-channel head coil at the Institute of Nuclear Sciences Applied to Health, Coimbra. First, a high-resolution (1 mm isovoxel) T1-weighted anatomical reference image was acquired from each participant using MPRAGE sequence. Four identical functional MR measurements were performed using Simultaneous MultiSlice imaging (66 slices), with six simultaneous slices, a flip angle of 68 degrees, an Echo Time of 37 ms, and a Repetition Time of 1000 ms. The functional data voxel was 2 mm (isovoxel).

2.3 Decoding pipeline

Acquired functional data were pre-processed using fMRIPrep [1]. The pipeline included slice timing correction, motion correction, susceptibility distortion correction using fieldmap images, registration of fMRI data from subject space to template MNI, and estimation and regression of confound signals. The data was converted to z-scores. In this sense, for each excerpt, a whole-brain image was obtained (representing the activation level of each voxel for that stimulus) (see Figure 2).

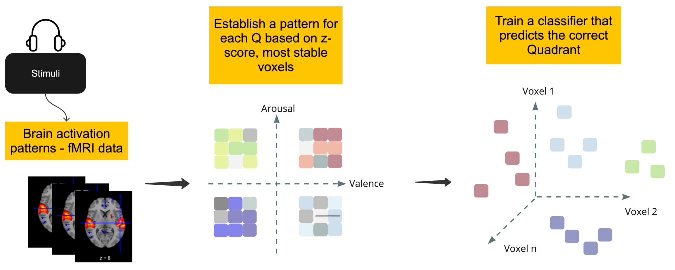

Figure 2: Decoding pipeline overview

The target (classification classes) was defined according to the labels provided by Panda et al. [3].

Statistical significance for each participant was assessed using permutation tests. The significance threshold was set at $p < 0.05$.

2.3.1 Feature selection

Multivariate pattern analysis involves defining features that characterise and allow the discrimination of the different classes in our classification framework, the classification algorithm, and assessment measures.

Given the duration of the stimuli (each musical excerpt lasts 11.5 seconds), we were interested in maximising the emotional response to the stimulus. To this end, to create each feature vector we removed the first 6 seconds. The remaining voxel activation levels data points were averaged (i.e.for each excertp, we created a single data point).

Regarding the spatial definition of the feature set, we estimated individual voxel stability masks. The most stable voxels in the brain are defined as voxels presenting a stable activation profile across the multiple presentations of a set of labels - defined here as the pairwise correlation.

2.3.2 Classification

Linear-kernel support vector machine classifiers were trained for each participant (one model for each participant) to perform the classification task, i.e., the prediction of the four quadrants Q1, Q2, Q3 and Q4 of Russell's circumplex. The classification was performed using the scikit-learn library. Each model was trained in a training set (70% of the data of each participant) to find the optimal number of voxels to use in the stability mask and to optimize the margin-parameter, C, in a grid-search 5-fold cross validation. After training, the optimal parameter set was used to predict labels on test set (30% of the data).

3 Results

3.1 Stability masks

The first step of the decoding pipeline was to identify the most stable voxels. To better understand the spatial distribution of the subset of voxels selected (ultimately used to create the feature vector and input to the classifier), we present the mask obtained combining the most stable voxels. The resulting mask, with a sparse distribution of voxels, is shown in Figure 3.

3.2 Decoding analysis

The average decoding accuracy obtained using the proposed pipeline was 62% ±15% (chance level in this multiclass problem is 25%). Statistical assessment based on permutation tests revealed that the decoding accuracy was significantly above chance level ($p < 0.05$) for 80% of the participants.

Figure 3: Feature selection - most stable voxels mask

To further characterise the prediction pattern of this model, we present the normalized confusion matrices associated to the predicted labels (the diagonal represents the percentage of correct predictions) (see Figure 4).

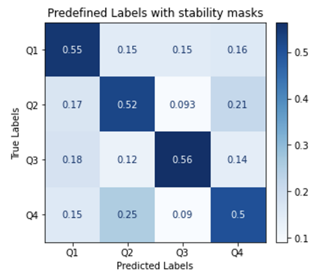

Figure 4: Confusion matrix normalized by the number of samples

The results show that the model tend to correctly identify the target label and no pattern emerges regarding the false negative.

4 Discussion

Constraining the total number of voxels based on a stability mask (feature selection) and validating our model (optimized for each participant) in an independent sample, we aimed at predicting the emotional quadrants of music stimuli based on fMRI data.

Significant voxel to discriminate valence and arousal were found in the multiple brain structures (in accordance with [2]). Clusters were found in frontal, parietal, temporal, occipital lobes, cingulate cortex, and cerebellum, noticeably in auditory cortex (superior temporal gyrus, Heschl's gyrus), primary somatosensory cortex (postcentral gyrus), and primary motor cortex (precentral gyrus), as well as supplementary and premotor cortices. Regarding clusters in the auditory cortex, we cannot rule out that these were based by acoustic differences between stimuli. Future work should include a control condition to account for the contribution of acoustical differences between stimuli and explore the contribution of acoustic, musical features of each excerpt to the evoked emotions and classification model performance.

Ultimately, the ability to discriminate valence and arousal of music stimuli, supporting the modulation of psychological and mood factors, will guide feedback in music-based interventions and contribute to neurofeedback learning outcomes.

References

[1] Oscar Esteban and et al. FMRIPrep: a robust preprocessing pipeline for functional MRI. *bioRxiv*, 2018. doi: 10.1101/306951.

[2] Stefan Koelsch. A coordinate-based meta-analysis of music-evoked emotions. *NeuroImage*, 223:117350, December 2020.

[3] Renato Panda, Ricardo Malheiro, and Rui Pedro Paiva. Novel Audio Features for Music Emotion Recognition. *IEEE Transactions on Affective Computing*, 11(4):614–626, October 2020.

[4] James A. Russell. A circumplex model of affect. *Journal of Personality and Social Psychology*, 39(6):1161–1178, 1980.

Acknowledgements

This work has been supported by Fundação para a Ciência e Tecnologia, grant EXPL/PSI-GER/0948/2021. Renato Panda was supported by Ci2 - FCT UIDP/05567/2020.